

УДК 004.522:004.031.6

Мартинюк І.А., Лахно В.А.  
Європейський університет, м. Київ**Структура та організація бази даних слів на основі файлової системи FAT в мікроконтролерних системах автоматичного розпізнавання мови**

Системи автоматичного розпізнавання мови складають основу голосного інтерфейсу інтерактивних інформаційних системи. В більшості випадків такі системи реалізуються у вигляді програмних додатків для персональних комп'ютерів чи серверів, оскільки потребують значних обчислювальних ресурсів. Проте, така реалізація накладає ряд обмежень, зокрема щодо портативності та автономності пристроїв, які використовують дану технологію. Вбудовані системи автоматичного розпізнавання мають ряд переваг порівняно з традиційними рішеннями та дозволяють ефективно вирішити задачу розпізнавання у таких областях, як автотранспорт, авіація, соціальна сфера, робототехніка тощо. Системи розпізнавання мови на базі сучасних високопродуктивних RISC-мікроконтролерів надають нові перспективи розвитку в даному напрямку завдяки підвищеному рівні відмовостійкості, високому рівні інтеграції, низькому рівні енергоспоживання та невисокій собівартості.

*Постановка проблеми.* Реалізація мікроконтролерних систем розпізнавання мови потребує наявності правильно організованої та структурованої бази даних, а також системи управління цією базою (СУБД). Для командних систем достатньо зберегти в пам'яті еталонні звукові записи та в процесі розпізнавання порівнювати звуковий сигнал з кожним таких записів. Таку базу даних просто організувати, проте подібні системи мають ряд недоліків, зокрема щодо обмеження кількості слів (команд), які можна розпізнати в реальному часі. Крім того, навіть при незначному збільшенні бази слів, в таких системах значно погіршується якість розпізнавання. Тому, для ефективного вирішення задачі розпізнавання, необхідно використовувати методи базовані на розпізнаванні частин мови (наприклад, ПММ) та подальшому визначенню їх ймовірних послідовностей у вигляді слів. Задачею бази даних, в даному випадку, буде зберігання набору слів, а основною задачею СУБД – вибірка та видача цих даних.

*Викладення основного матеріалу.* Для побудови такої СУБД потрібно визначитись із фізичним розміщенням бази даних. В простому випадку, можна записувати дані безпосередньо у власну flash-пам'ять мікроконтролера. Проте, розміру такої пам'яті не достатньо для збереження та правильної організації бази, тому слід задіяти зовнішню пам'ять. Найперспективнішим рішенням, в даному випадку, є використання Secure Digital (SD) карти пам'яті, що дозволить зберігати значні об'єми даних. А завдяки поширеним інтерфейсам зв'язку (SPI та SDIO), які підтримують даний тип карт, апаратно з'єднати їх із мікроконтролером.

В процесі розпізнавання, часто виникають ситуації, коли однозначно не можливо визначити, яка частина мови (спрощено, будемо вважати за літеру алфавіту) представлена в даному фреймі [1]. При цьому, в результаті порівняння звукового сигналу фрейму із базою еталонних записів, відомі ймовірнісні характеристики появи таких частин (рис.1).

Може скластись враження, що для того щоб розпізнати слово із звукового сигналу, достатньо побудувати його із послідовності літер з максимальними ймовірнісними характеристики. Проте така спроба, в більшості випадків, буде не вдалою, або навіть не коректною, оскільки такого слова може не бути в словнику відповідної мови. Тому, для задачі розпізнавання безперервного мовлення з великим розміром словнику, необхідна структура даних, яка буде об'єднувати всі реально можливі варіанти розпізнавання. Найбільш вдалим варіантом такої структури є скінченний автомат з варіантами розпізнавання та їх ймовірнісними



характеристиками – WFST (Weighted Finite-State Transducers) [2], яка забезпечує загальне представлення про ПММ, граматику та альтернативні варіанти розпізнавання (рис. 2). При цьому, процес розпізнавання зводиться до вибору кращого шляху в даному графі [3].

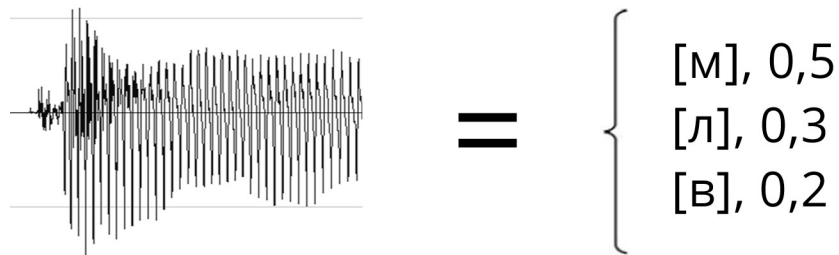


Рис. 1. Ймовірності появи частин мови в фреймі

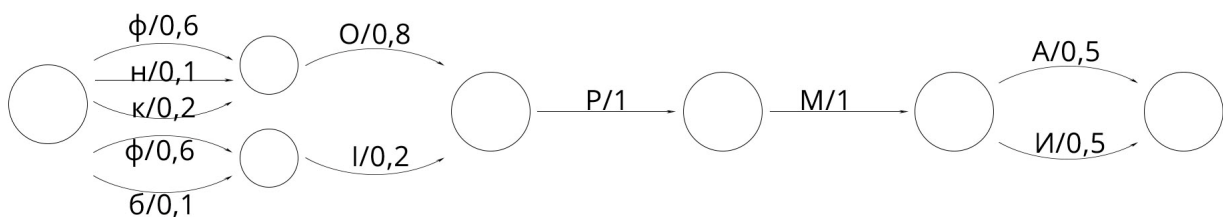


Рис. 2. Приклад WFST графа

В якості бази даних слів, як складової частини такої структури, може слугувати файлова система FAT, оскільки вона являє собою просту реалізацію графа. Тоді ребрами графу слугуватимуть директорії, кожна з яких буде відображати літеру в конкретній позиції слова. Вкладеність таких директорій складатиме абсолютних шлях у файловій системі, що відповідає конкретному слову, наприклад:

- \\ф\i\р\м\а\;
- \\ф\o\р\м\а\;
- \\н\o\р\м\а\.

Вага ребер графу, в даному випадку, визначатиметься ще на етапі розпізнавання фрейму (рис.1). Тоді для визначення слова, як кінцевого результату розпізнавання, необхідно пройти наступні кроки:

- 1) визначити всі максимально можливі комбінації літер відповідно ймовірнісних характеристик;
- 2) відкинути із набору неіснуючі варіанти слів, перевіривши на існування кожного шляху в файловій системі;
- 3) визначити найбільш відповідний результат розпізнавання, шляхом обчислення середнього арифметичного значення його ймовірностей.

**Висновки.** Обґрунтовано необхідність організації бази даних слів для систем автоматичного розпізнавання мови та запропоновано структуру такої бази на основі файлової системи FAT. Перевагою такого рішення є простота реалізації та швидкість вибірки інформації з бази. Серед недоліків слід відзначити обмеження файлової системи відносно максимальної кількості директорій та дещо завищені витрати ресурсів для підтримки кирилиці.

#### Список використаних джерел

1. Varshney N. Embedded speech recognition system [Text] / N. Varshney, S. Singh // International journal of advanced research in electrical, electronics and instrumentation energy. – 2014. – Vol. 3, № 4. – P. 9218- 9227.
2. Mohri M. Weighted Finite-State Transducers in Speech Recognition. [Text] / M. Mohri, F. Pereira, M. Riley // Computer Speech & Language. – 2002. – Vol. 16, Issue 1. – P. 69-88.
3. Распознавание речи от Яндекса. Под капотом у Yandex.SpeechKit [Электрон. ресурс]. – Режим доступа: <https://habrahabr.ru/company/yandex/blog/198556/>.